Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal Pre-proofs

Research Article

Origin and evolutionary analysis of the SARS-CoV-2 Omicron variant

Yamin Sun, Wenchao Lin, Wei Dong, Jianguo Xu

PII: S2588-9338(21)00055-8
DOI: https://doi.org/10.1016/j.jobb.2021.12.001
Reference: JOBB 92

To appear in: *Journal of Biosafety and Biosecurity*

Received Date: 10 December 2021
Revised Date: 16 December 2021
Accepted Date: 19 December 2021

Please cite this article as: Y. Sun, W. Lin, W. Dong, J. Xu, Origin and evolutionary analysis of the SARS-CoV-2 Omicron variant, *Journal of Biosafety and Biosecurity* (2021), doi: https://doi.org/10.1016/j.jobb.2021.12.001

Origin and evolutionary analysis of the SARS-CoV-2 Omicron variant

Yamin Sun[1,a,b], Wenchao Lin[1,b], Wei Dong[b], Jianguo Xu[a,c,*]

[a] Research Institute of Public Health, Nankai University, Tianjin, PR China
[b] Research Center for Functional Genomics and Biochip, Tianjin, PR China
[c] State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, PR China

* Corresponding author. Jianguo Xu (xujianguo@icdc.cn)
[1] These authors contributed equally to this work.

ABSTRACT

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has evolved rapidly into new variants throughout the pandemic. The Omicron variant has more than 50 mutations when compared with the original wild-type strain and has been identified globally in numerous countries. In this report, we analyzed the mutational profiles of several variants, including the per-site mutation rate, to determine evolutionary relationships. The Omicron variant was found to have a unique mutation profile when compared with that of other SARS-CoV-2 variants, containing mutations that are rare in clinical samples. Moreover, the presence of five mouse-adapted mutation sites suggests that Omicron may have evolved in a mouse host. Mutations in the Omicron receptor-binding domain (RBD) region, in particular, have potential implications for the ongoing pandemic.

Keywords: SARS-CoV-2, Omicron variant, mouse-adapted mutation, reverse zoonosis

1. Introduction

The current COVID-19 pandemic is a global human health crisis caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). On 26 November 2021, the World Health Organization (WHO) designated the SARS-CoV-2 variant B.1.1.529, named Omicron, as its fifth variant of concern (VOC). This decision was based on the evidence presented to health officials and researchers that Omicron had numerous mutations with potential implications for the ongoing pandemic. The Omicron variant has now been identified globally[1], including countries throughout Asia, Africa, Europe, and North America.

The original wild-type SARS-CoV-2 strain likely originated in a bat host[2-4]. Initially, pangolins were thought to be the source of spillover to humans, but they may have been infected by other animal species[5]. Since the outbreak of COVID-19, several countries have reported infections of SARS-CoV-2 in animals. Human-to-animal transmission has been observed in pets, farmed animals, and animals held in zoos, in addition to free-ranging wild animals[6,7]. For example, infections under natural conditions have been reported in pet dogs[8] and cats[9], in farmed mink[10] and ferrets[11], and tigers, lions, snow leopards, pumas, and gorillas at zoos[12]. Most diseased animals are hypothesized to have been infected through close contact with COVID-19-positive human patients. However, no compelling evidence currently shows that any domestic animal can readily transmit SARS-CoV-2 to other animals, including humans. Few animal cases have shown the potential for further zoonotic and anthroponotic viral transmission. Nevertheless, infection in domestic and wild animal species has possible implications for public health.

SARS-CoV-2 enters host cells via the interaction of spike-like proteins (S proteins) on the viral surface with the host cell entry receptor angiotensin-converting enzyme 2 (ACE2)[2]. Some variants that have mutations in the receptor-binding domain (RBD) region of the S protein are VOC because they are potentially associated with enhanced transmission, pathogenicity, and/or immune evasion[13]. Although the initial wild-type strain of SARS-CoV-2 does not infect mice, mouse-adapted SARS-CoV-2 strains have been identified. Several mouse-adapted strains have mutations located in the RBD region, enhancing interactive affinities with mouse ACE2 (mACE2)[14] to facilitate efficient viral replication in this host. A mouse-adapted strain at passage 6 (MASCp6), which has an N501Y mutation, was shown to have increased infectivity in the lung during serial passaging in BALB/c mice[15]. Another study showed that three SARS-CoV-2 VOCs, namely B.1.1.7 and two other N501Y-carrying variants, B.1.351 and P.3, can infect mice[16].

In this study, we constructed a phylogenetic tree of all known VOCs and variants of interest (VOIs). The results showed that the Omicron variant was not present on an intermediate evolutionary branch, suggesting that it may have evolved in a non-human host. Analysis of Omicron mutation data revealed a high number of mutations, that these mutations are concentrated in the S protein (specifically the

RDB region), and that Omicron has five mouse-adapted mutation sites. Together, the data suggest that Omicron may have evolved in a mouse host.

2. Materials and methods

2.1. Data collection

We downloaded a representative set of SARS-CoV-2 genomes from individuals infected during the COVID-19 pandemic from the GISAID database[17]. The genomes had complete metadata, including patient age and sex and the year and country in which samples were collected. These data were used to test associations between variation in SARS-CoV-2 genomes and available epidemiological metadata.

2.2. Mutation analysis

The complete genome of SARS-CoV-2 isolate Wuhan-Hu-1 (NC_045512.2) was used as the reference genome[18], and mutations in all other samples were compared with this reference isolate. Detected mutations were confirmed with Integrative Genomics Viewer (IGV) and annotated with the SnpEff program[19].

2.3. Construction of a phylogenetic tree with full-length genomic sequences

The full-length genomic sequences of VOCs and VOIs used in this analysis included 30 each of the Alpha, Beta, Eta, Iota, Mu, Kappa, Zeta, Theta, Epsilon, and Omicron variants. There were also 28 Gamma, 98 Delta, and 29 Lambda variant genomes included. All 455 genomes were aligned using MAFFT v7.31023[20]. The aligned sequences were converted to the phylip file format with Clustal W[21], and maximum likelihood (ML) trees were then constructed in RaxML v8.2.12[22] with 100 bootstrap replicates. The time-scaled phylogenetic trees were constructed using NextStrain[23] and Treetime[24] and visualized with FigTree v1.4.4[25].

3. Results and discussion

3.1. High number of mutations

We calculated the average number of mutations in the five VOCs circulating globally and found that the Omicron variant has significantly more mutations than any other variant currently in circulation (Table 1). This observation suggests that the environment in which Omicron evolved may differ from other known VOCs that have infected healthy human hosts. The Omicron variant likely evolved in an immunocompromised patient, although it is possible that this variant also evolved in an animal host.

3.2. Key mutation positions

The RBD region recognizes ACE2, the host receptor that binds to the viral S

protein[26]. Mutations in the RBD region may increase the binding affinity and viral infectivity. Furthermore, most vaccine-induced neutralizing antibodies and antibody treatments target the RBD. The Omicron variant has at least 15 mutations in the RBD region, including mutations at Q493 and Q498 (Fig. 1), which are especially concerning to public health experts. Studies have shown that mutations at these two sites are related to the infectivity of animals. In 2021, the Jin research team showed that strains with the Q493K and Q498H mutations have significantly enhanced affinity toward mACE2[14]. In a study of New York sewer samples published in July 2021[27], researchers found many variants with the Q493K and Q498Y mutations, which were rare in clinical samples. At that time, only three reported strains of SARS-CoV-2 had the Q498H mutation, and none had the Q498Y mutation. This study showed that by July 2021, the Q498 mutation had accumulated in large numbers of animal hosts living in the sewers of New York, and the authors discussed the possibility of SARS-CoV-2 spreading between non-human animal hosts. A CSIRO study additionally identified seven key mutation sites potentially related to mACE2 binding affinity. In the S protein, these sites are K417, E484, F486, Q493, Q498, P499, and N501[28-31]. We compared key mutations in 13 mouse-adapted strains with the Omicron variant (Fig. 2). The results showed that the Omicron variant contains mutations at five key sites of viral S protein: K417, E484, Q493, Q498, and N501. Notably, another strain had mutations at the same five sites, the IA-501Y-MA-30 strain, which was obtained from mouse lung samples after 30 passages of the IA-501Y strain[32]. These results suggest that the Omicron variant may have evolved in a mouse host.

3.3. Phylogenetic analysis of VOCs and VOIs

Despite a large number of mutations in Omicron, no evidence was found in known public databases to suggest that these mutations slowly accumulated over time. Additionally, phylogenetic trees showed no intermediate branches of evolution, which is a very surprising result. Starting in August 2021, the Delta variant was dominant globally, and until November 2021, 99.6% of all collected specimens causing new infections were identified as Delta (Fig. 3A). If Omicron evolved from a strain of the Delta variant, such as AY.4, AY.23, or AY.46 (the dominant variants in Europe, Asia, and Africa, respectively), they would share a common mutation profile. However, analysis of data from GISAID showed that the Omicron variant differed from each of these strains and did not evolve from the Delta variant (Fig. 3B). The phylogenetic analysis strongly indicates that the Omicron variant forms a monophyletic group with the Gamma variant as a sister group, and the Omicron group has an extremely long branch length. The time-scaled phylogenetic tree shows that the Omicron and Gamma lineages likely diverged in the first half of 2020. This supports the hypothesis that Omicron may have evolved in a non-human animal species. After accumulating many mutations in the animal host, the altered coronavirus was transmitted back to humans

by reverse zoonosis.

The emergence of the Omicron variant indicates that surveillance of SARS-CoV-2 variants should be conducted in economically underdeveloped countries and in the environment to avoid the continuous emergence of new variants of unknown origin. Understanding the threat posed by the Omicron variant will require researchers to gather and analyze a great deal more data in a brief period. Determining the origin of Omicron requires surveillance of animals, especially rodents, because they may have come into contact with humans carrying a strain of the virus with adaptive mutations. Future work should focus on SARS-CoV-2 variants isolated from other wild animals to investigate the evolutionary trajectories and biological properties of these variants both *in vitro* and *in vivo*. If Omicron is determined to have been derived from animals, the implications of it circulating among non-human hosts will pose new challenges in the prevention and control of the epidemic.

References

1. Karim SSA, Karim QA. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. The Lancet. 2021;398(10317):2126-2128. doi:10.1016/S0140-6736(21)02758-6
2. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020/03/01 2020;579(7798):270-273. doi:10.1038/s41586-020-2012-7
3. Lau SK, Woo PC, Li KS, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A. Sep 27 2005;102(39):14040-5. doi:10.1073/pnas.0506735102
4. Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. Nature Reviews Microbiology. 2019/03/01 2019;17(3):181-192. doi:10.1038/s41579-018-0118-9
5. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet. 2020;395(10224):565-574. doi:10.1016/S0140-6736(20)30251-8
6. Authority EFS, Prevention ECfD, Control, et al. Monitoring of SARS-CoV-2 infection in mustelids. EFSA Journal. 2021;19(3):e06459. doi:https://doi.org/10.2903/j.efsa.2021.6459
7. Korath ADJ, Janda J, Untersmayr E, et al. One Health: EAACI Position Paper on coronaviruses at the human-animal interface, with a specific focus on comparative and zoonotic aspects of SARS-Cov-2. Allergy. n/a(n/a)doi:https://doi.org/10.1111/all.14991
8. Morgan L, Protopopova A, Birkler RID, et al. Human–dog relationships during the COVID-19 pandemic: booming dog adoption during social isolation. Humanities and Social Sciences Communications. 2020/11/24 2020;7(1):155. doi:10.1057/s41599-020-00649-x
9. Sharun K, Saied AA, Tiwari R, Dhama K. SARS-CoV-2 infection in domestic and feral cats: current evidence and implications. Vet Q. 2021;41(1):228-231. doi:10.1080/01652176.2021.1962576
10. Rabalski L, Kosinski M, Smura T, et al. Severe Acute Respiratory Syndrome Coronavirus 2 in Farmed Mink (Neovison vison), Poland. Emerging Infectious Disease

journal. 2021;27(9):2333. doi:10.3201/eid2709.210286

11. Kim YI, Kim SG, Kim SM, et al. Infection and Rapid Transmission of SARS-CoV-2 in Ferrets. Cell Host Microbe. May 13 2020;27(5):704-709.e2. doi:10.1016/j.chom.2020.03.023
12. Mathavarajah S, Dellaire G. Lions, tigers and kittens too: ACE2 and susceptibility to COVID-19. Evol Med Public Health. 2020;2020(1):109-113. doi:10.1093/emph/eoaa021
13. Dinnon KH, 3rd, Leist SR, Schäfer A, et al. A mouse-adapted model of SARS-CoV-2 to test COVID-19 countermeasures. Nature. Oct 2020;586(7830):560-566. doi:10.1038/s41586-020-2708-8
14. Huang K, Zhang Y, Hui X, et al. Q493K and Q498H substitutions in Spike promote adaptation of SARS-CoV-2 in mice. EBioMedicine. 2021;67doi:10.1016/j.ebiom.2021.103381
15. Gu H, Chen Q, Yang G, et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. Science. Sep 25 2020;369(6511):1603-1607. doi:10.1126/science.abc4730
16. Shuai H, Chan JF-W, Yuen TT-T, et al. Emerging SARS-CoV-2 variants expand species tropism to murines. EBioMedicine. 2021;73doi:10.1016/j.ebiom.2021.103643
17. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. Mar 30 2017;22(13)doi:10.2807/1560-7917.Es.2017.22.13.30494
18. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020/03/01 2020;579(7798):265-269. doi:10.1038/s41586-020-2008-3
19. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). Apr-Jun 2012;6(2):80-92. doi:10.4161/fly.19695
20. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution. 2013;30(4):772-780. doi:10.1093/molbev/mst010
21. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. Nov 11 1994;22(22):4673-80. doi:10.1093/nar/22.22.4673
22. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-1313. doi:10.1093/bioinformatics/btu033
23. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34(23):4121-4123. doi:10.1093/bioinformatics/bty407
24. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evolution. 2018;4(1)doi:10.1093/ve/vex042
25. Rambaut A. FigTree. http://tree.bio.ed.ac.uk/software/figtree/
26. Wang Q, Zhang Y, Wu L, et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell. 2020/05/14/ 2020;181(4):894-904.e9. doi:https://doi.org/10.1016/j.cell.2020.03.045
27. Trujillo M, Cheung K, Gao A, et al. Protocol for safe, affordable, and reproducible

isolation and quantitation of SARS-CoV-2 RNA from wastewater. PLOS ONE. 2021;16(9):e0257454. doi:10.1371/journal.pone.0257454

28. Verkhivker GM, Agajanian S, Oztas DY, Gupta G. Comparative Perturbation-Based Modeling of the SARS-CoV-2 Spike Protein Binding with Host Receptor and Neutralizing Antibodies: Structurally Adaptable Allosteric Communication Hotspots Define Spike Sites Targeted by Global Circulating Mutations. Biochemistry. 2021;60(19):1459-1484. doi:10.1021/acs.biochem.1c00139
29. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. May 2020;581(7807):215-220. doi:10.1038/s41586-020-2180-5
30. Muruato A, Vu MN, Johnson BA, et al. Mouse Adapted SARS-CoV-2 protects animals from lethal SARS-CoV challenge. bioRxiv. May 4 2021;doi:10.1101/2021.05.03.442357
31. Kuiper MJ, Wilson LO, Mangalaganesh S, Lee C, Reti D, Vasan SS. But Mouse, you are not alone: On some severe acute respiratory syndrome coronavirus 2 variants infecting mice. bioRxiv. 2021:2021.08.04.455042. doi:10.1101/2021.08.04.455042
32. Roy Wong LY, Zheng J, Wilhelmsen K, et al. Eicosanoid signaling as a therapeutic target in middle-aged mice with severe COVID-19. bioRxiv. Apr 21 2021;doi:10.1101/2021.04.20.440676

Figure captions

Fig. 1. Mutation profiles of the five variants of concern (VOCs) designated by the World Health Organization. Common mutations are marked in red. Many mutations in the Omicron variant are unique when compared with mutations in other VOCs.

Fig. 2. Seven key mutations among 13 mouse-adapted strains and the Omicron variant of SARS-CoV-2. IA-501Y-MA-30 is a homogenate obtained from mouse lung after 30 passages of the IA-501Y strain.

Fig. 3. The abundance of different SARS-Cov-2 variants by percent of new infections from a public database (A). Evolutionary analysis shows that the Omicron variant did not evolve from the Delta variant and differs from other variants (B).

Tables

Table 1
The average number of mutations per sample in each VOC is based on publicly available datasets.

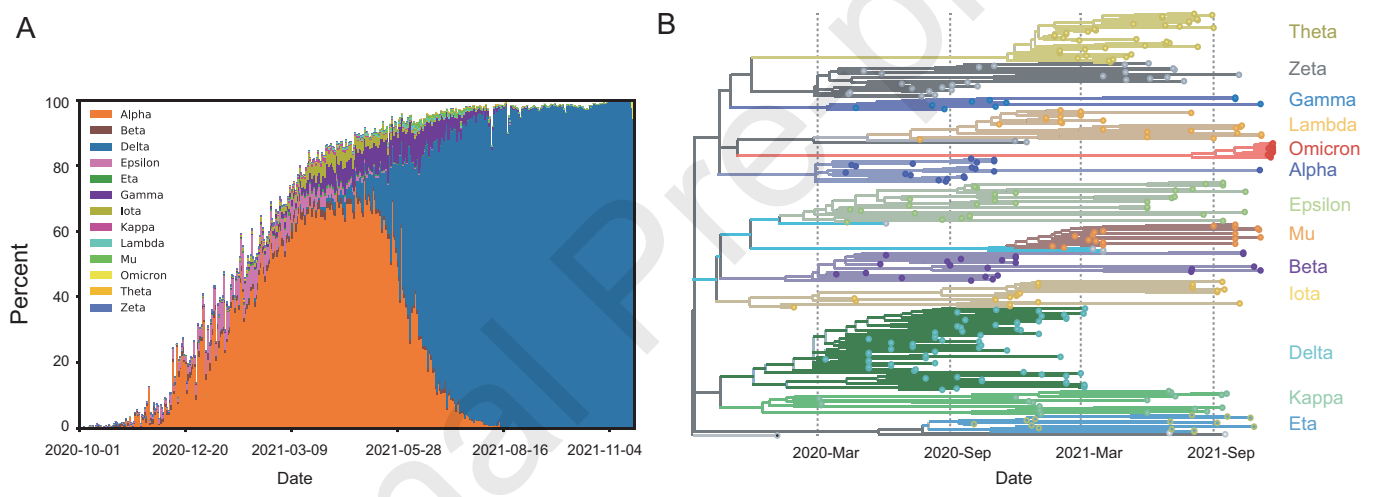
Variant	Average number of mutations
Alpha	29.7
Gamma	29.1
Beta	28.4
Delta	35.4
Omicron	53.3

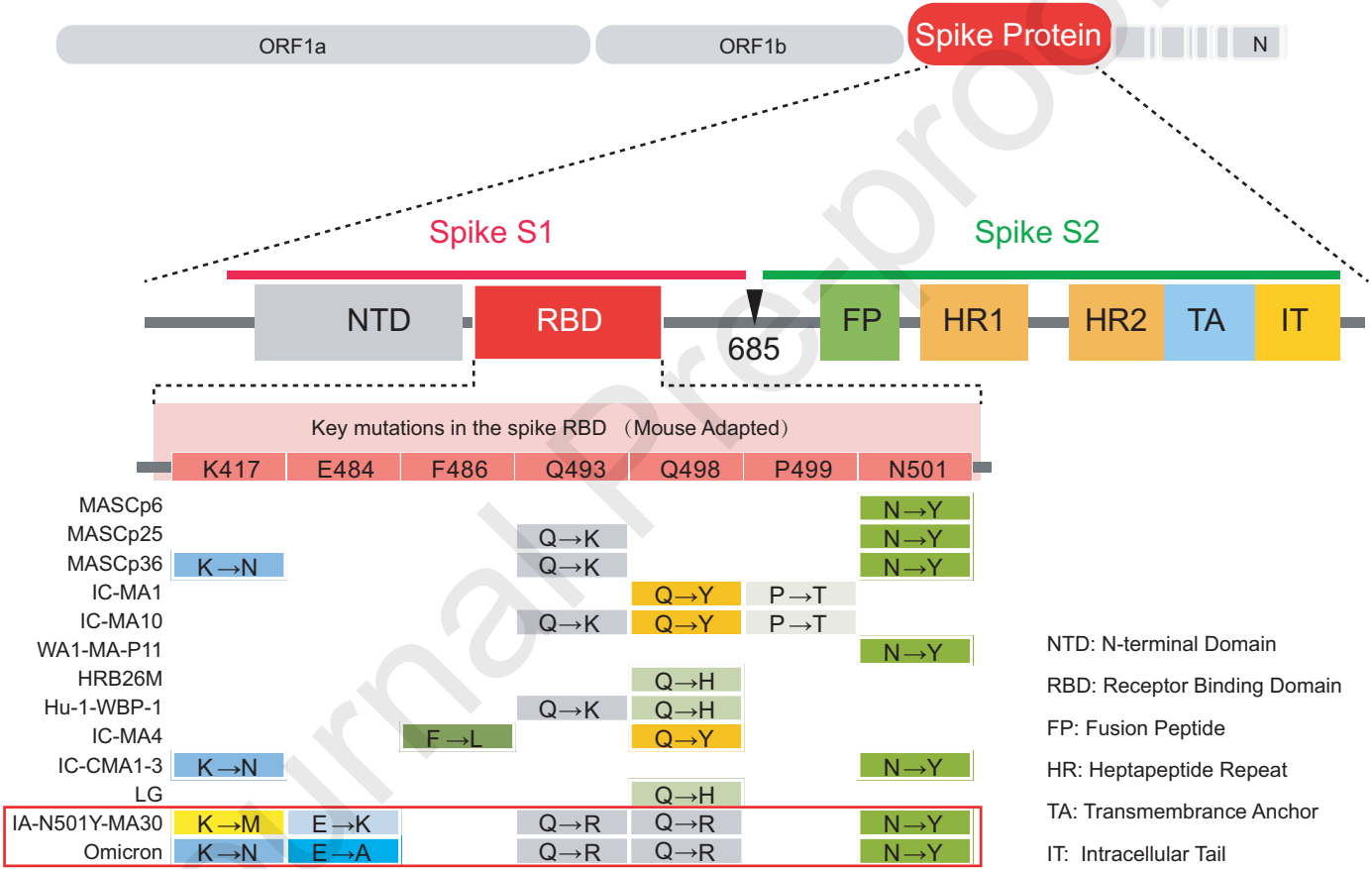
CRediT authorship contribution statement:

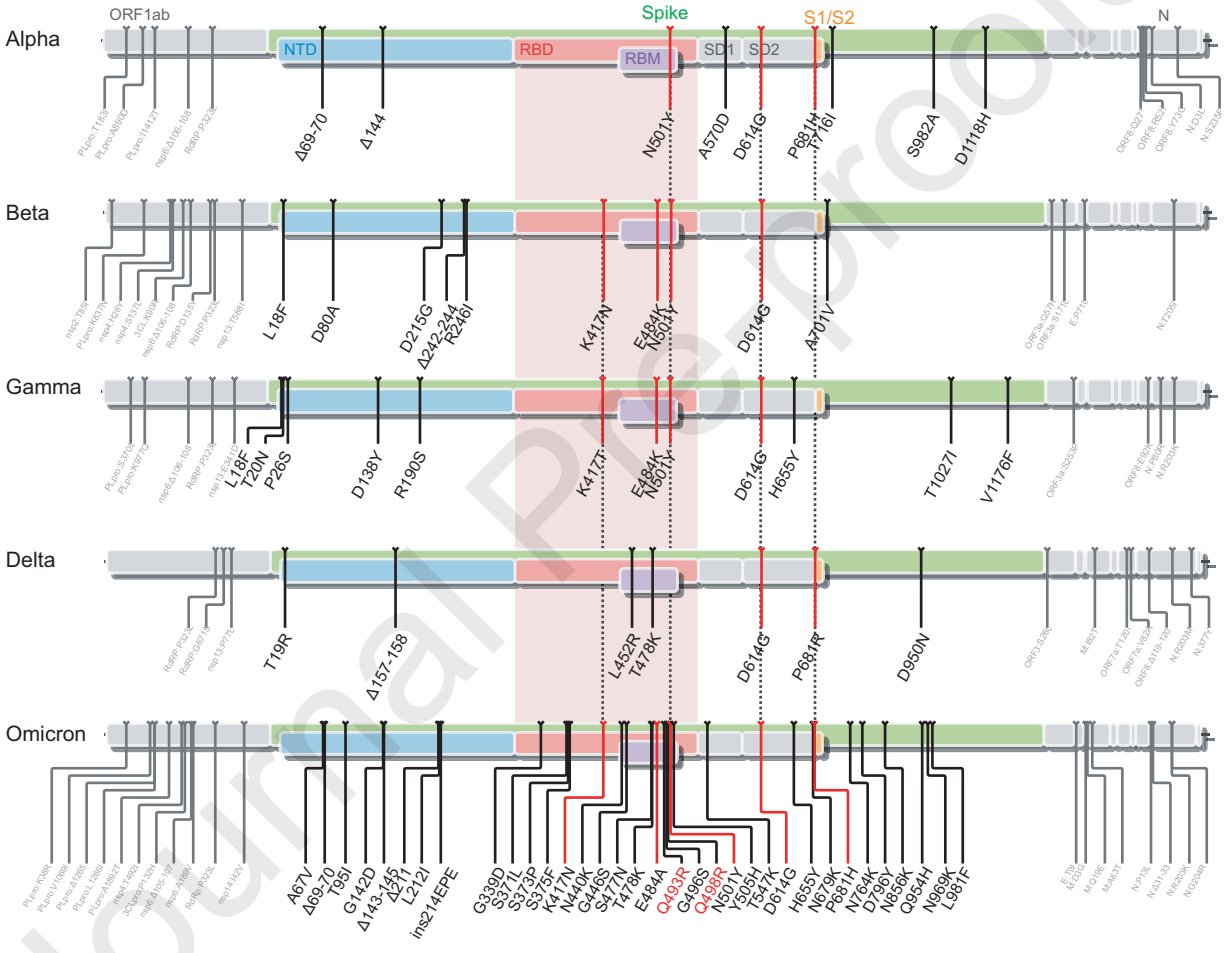
Yamin Sun: Writing - original draft, Visualization. **Wenchao Lin**: Writing - original draft. **Wei Dong**: Software, Data curation. **Jianguo Xu**: Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.


A
Percent
100
80
60
40
20
0
Alpha
Beta
Delta
Epsilon
Eta
Gamma
Iota
Kappa
Lambda
Mu
Omicron
Theta
Zeta
2020-10-01
2020-12-20
2021-03-09
2021-05-28
2021-08-16
2021-11-04
Date
B
Theta
Zeta
Gamma
Lambda
Omicron
Alpha
Epsilon
Mu
Beta
Iota
Delta
Kappa
Eta
2020-Mar
2020-Sep
2021-Mar
2021-Sep
Date


ORF1a
ORF1b
Spike Protein
N
Spike S1
Spike S2
NTD
RBD
685
FP
HR1
HR2
TA
IT
Key mutations in the spike RBD （Mouse Adapted）
K417
E484
F486
Q493
Q498
P499
N501
MASCp6
MASCp25
MASCp36
IC-MA1
IC-MA10
WA1-MA-P11
HRB26M
Hu-1-WBP-1
IC-MA4
IC-CMA1-3
LG
IA-N501Y-MA30
Omicron
K→N
Q→K
N→Y
Q→Y
P→T
Q→H
F→L
K→M
E→K
E→A
Q→R
NTD: N-terminal Domain
RBD: Receptor Binding Domain
FP: Fusion Peptide
HR: Heptapeptide Repeat
TA: Transmembrance Anchor
IT: Intracellular Tail


ORF1ab
Spike
S1/S2
N
Alpha
NTD
RBD
RBM
SD1
SD2
Δ69-70
Δ144
N501Y
A570D
D614G
P681H
T716I
S982A
D1118H
Beta
L18F
D80A
D215G
Δ242-244
R246I
K417N
E484K
N501Y
D614G
A701V
Gamma
L18F
T20N
P26S
D138Y
R190S
K417T
E484K
N501Y
D614G
H655Y
T1027I
V1176F
Delta
T19R
Δ157-158
L452R
T478K
D614G
P681R
D950N
Omicron
A67V
Δ69-70
T95I
G142D
Δ143-145
Δ211
L212I
ins214EPE
G339D
S371L
S373P
S375F
K417N
N440K
G446S
S477N
T478K
E484A
Q493R
G496S
Q498R
N501Y
Y505H
T547K
D614G
H655Y
N679K
P681H
N764K
D796Y
N856K
Q954H
N969K
L981F